Q. J. R. Meteorol. Soc. (2000), **126**, pp. 2933–2949

Multiresolution analysis, entropic information and the performance of atmospheric sounding radiometers

By G. E. PECKHAM*

Heriot-Watt University, Edinburgh

April 2000

SUMMARY

The performance of remote sounding radiometers measuring properties of the Earth's atmosphere is analysed through a multiresolution wavelet transform. This technique allows the uncertainty in retrieved atmospheric profiles to be determined as a function of both altitude and scale of patterns in the profile. Multiresolution analysis may be applied to a number of indicators of the quality of a measurement including entropic information. The apportionment of performance indicators to specific altitude ranges and pattern scales facilitates a comparison with performance requirements. The analysis is illustrated through a simple model of a remote sounding radiometer and by application to the Infrared Atmospheric Sounding Interferometer.

KEYWORDS: Atmosphere Entropy Information theory Infrared Atmospheric Sounding Interferometer (IASI) Multi-resolution Radiometry Remote sensing Remote sounding Retrieval Wavelet

1. INTRODUCTION

Temperature and composition profiles of the Earth's atmosphere have been measured almost continuously by a succession of space-borne radiometers following Kaplan's (1959) description of this technique. Similar instruments have been mounted on aircraft, ground platforms and planetary probes (Houghton and Taylor 1973; Houghton *et al.* 1984). Remote measurements from spacecraft have the advantage of continuous global coverage so that they have become an established part of the world weather monitoring system in spite of their poor vertical resolution compared with in-situ measurements. A new generation of instruments with higher spectral resolution and wider spectral coverage has been designed to provide improved vertical resolution (Smith *et al.* 1990; Aumann and Pagano 1994; Simeoni *et al.* 1997).

Although it has been recognised since the work of Backus and Gilbert (1970) that the achievable resolution must be balanced against noise in the retrieved profile, there appears to be no generally agreed definition of resolution. This leads to difficulty in defining and interpreting performance requirements. An unambiguous measure of performance is the probability of error in detecting prescribed structures in the atmospheric profile in the presence of measurement noise (Prunet *et al.* 1998; Lee 1998b,c). Global measures such as the degrees of freedom for signal (Wahba 1985) and Shannon's entropic information (Shannon 1949) give no detail about the distribution of performance over altitude and scale. Purser and Huang (1993) addressed this problem by deriving an effective data density over altitude and later a density of entropic information (Huang and Purser 1996). Here I show that, by means of a multiresolution wavelet transform, global measures of performance may be distributed over both altitude and scale of pattern in the atmospheric profile. Performance requirements match naturally to such a decomposition.

* Corresponding author: Department of Physics, Heriot-Watt University, Riccarton, Edinburgh EH14 4AS, UK. e-mail: g.e.peckham@hw.ac.uk

2. Retrieval of the atmospheric state

A remote sensing radiometer measures the radiances emitted by the atmosphere in a number of spectral bands or channels. The vector $\mathbf{y}$ representing the radiances measured by N channels is related to the atmospheric state specified by the M dimensional vector $\mathbf{x}$ through a forward model $\mathbf{y} = \mathbf{f}(\mathbf{x}) + \mathbf{e}$. Here the vector $\mathbf{e}$ includes the errors of measurement and errors arising through uncertainties in the model. In a linear approximation about some state $\mathbf{x}^0$

$$\mathbf{y} = \mathbf{K}\mathbf{x} + \mathbf{e} \tag{1}$$

where $\mathbf{K}$ is the Jacobian matrix with elements $[\mathbf{K}]_{ij} = \partial \mathbf{f}_i / \partial \mathbf{x}_j$ evaluated at $\mathbf{x} = \mathbf{x}^0$ and $\mathbf{x}$ is now measured from $\mathbf{x}^0$ and $\mathbf{y}$ from $\mathbf{y}^0 = \mathbf{f}(\mathbf{x}^0)$. The atmospheric state vector includes all atmospheric parameters which are to be retrieved from the measurements. Typically $\mathbf{x}$ will include, as a minimum, atmospheric temperatures and water vapour concentrations tabulated over a range of altitudes (the components of $\mathbf{x}$ will be assumed to be in order of increasing altitude with composition profiles appended to the temperature profile). For simplicity in this paper I consider only temperatures, but the extension to water vapour concentrations is straightforward.

The statistical method for retrieving an estimate of the atmospheric state from the measurements $\mathbf{y}$ and a prior estimate $\mathbf{x}^{\mathrm{b}}$ of $\mathbf{x}$ has been described by many authors (e.g. Wark and Fleming 1966; Rodgers 1976; Menke 1984; Rodgers 2000). If all random vectors have multivariate normal distributions, the most probable value of $\mathbf{x}$ after the measurement is $\mathbf{x}^{\mathrm{a}}$ where

$$\mathbf{x}^{\mathrm{a}} - \mathbf{x}^{\mathrm{b}} = \mathbf{A}\mathbf{K}^{\mathrm{T}}\mathbf{E}^{-1}(\mathbf{y} - \mathbf{K}\mathbf{x}^{\mathrm{b}}) \tag{2}$$

$$\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{K}^{\mathrm{T}}\mathbf{E}^{-1}\mathbf{K} \tag{3}$$

Here $\mathbf{A}$ is the covariance[†] of $\mathbf{x}$ after the measurement, $\mathbf{E}$ is the covariance of the measurement and forward model error $\mathbf{e}$ and $\mathbf{B}$ is the covariance of $\mathbf{x}$ before the measurement. The post measurement probability distribution of $\mathbf{x}$ is

$$\mathrm{N}_M(\mathbf{x} - \mathbf{x}^{\mathrm{a}}, \mathbf{A}) = (2\pi)^{-M/2}|\mathbf{A}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{x}^{\mathrm{a}})^{\mathrm{T}}\mathbf{A}^{-1}(\mathbf{x} - \mathbf{x}^{\mathrm{a}})\right\} \tag{4}$$

where $|\mathbf{A}|$ denotes the determinant of $\mathbf{A}$. The measurement error distribution is $\mathrm{N}_N(\mathbf{e}, \mathbf{E})$ ($\mathbf{e}$ is unbiased) and the probability distribution of $\mathbf{x}$ before the measurement is $\mathrm{N}_M(\mathbf{x} - \mathbf{x}^{\mathrm{b}}, \mathbf{B})$. Equations 2 and 3 apply if the measurement errors are statistically independent of the prior estimate of the atmospheric state.

The retrieval process described above is referred to as assimilation when the prior estimate is from a numerical weather model and the measurements are being used to update the model. The statistical retrieval method gives consistent results when applied sequentially. The set of measurements may be divided into two statistically independent subsets and an intermediate estimate made by assimilating only the first subset. This intermediate estimate and its covariance may be used as the prior estimate for the assimilation of the second subset. The resulting final estimate and its covariance are identical to the results which would have been obtained by assimilating the complete set of measurements in one application of equations 2 and 3.

[†] Notation is summarised in appendix A. Vectors are denoted by bold lower case letters and matrices by bold upper case letters.

3. The resolution matrix

Many authors have used measures of resolution based on the resolution matrix $\mathbf{R}$. Conrath (1972) applied methods developed by Backus and Gilbert (1970) in the context of sounding the solid earth to the atmospheric sounding problem. The change in the estimated atmospheric profile $\mathbf{x}^{\mathrm{a}}$ caused by a perturbation $\mathbf{x}$ in the true profile is given in a linear approximation by

$$\langle \mathbf{x}^{\mathrm{a}} \rangle_{\mathbf{e}} = \mathbf{R}\,\mathbf{x} \tag{5}$$

where $\langle \cdots \rangle_{\mathbf{e}}$ is an ensemble average over measurement noise. The spread of significant elements in the rows of the matrix $\mathbf{R}$ determines the range of altitudes over which the true profile is averaged in the estimate $\mathbf{x}^{\mathrm{a}}$. The requirement that this spread should be small conflicts with the requirement that the variances of the retrieved values arising from measurement noise should also be small. A compromise may be made by choosing retrieval coefficients to minimise a weighted sum of the spread and the variance. By changing the weights, a trade-off curve may be plotted of variance vs. resolution.

In the statistical method outlined in section 2, this trade-off is determined by the relative values of prior and measurement covariance matrices and an unambiguous value for $\mathbf{R}$ is obtained. The measurement noise is supposed unbiased so that $\langle \mathbf{y} \rangle_{\mathbf{e}} = \mathbf{K}\mathbf{x}$. For a given prior estimate $\mathbf{x}^{\mathrm{b}}$

$$\langle \mathbf{x}^{\mathrm{a}} \rangle_{\mathbf{e}} = \mathbf{A}\,\mathbf{K}^{\mathrm{T}}\mathbf{E}^{-1}\mathbf{K}\,(\mathbf{x} - \mathbf{x}^{\mathrm{b}}) + \mathbf{x}^{\mathrm{b}} \tag{6}$$

$$= \mathbf{R}\mathbf{x} + (\mathbf{I} - \mathbf{R})\mathbf{x}^{\mathrm{b}} \tag{7}$$

where $\mathbf{I}$ is the unit matrix and

$$\mathbf{R} = \mathbf{A}\,\mathbf{K}^{\mathrm{T}}\mathbf{E}^{-1}\mathbf{K} = \left(\mathbf{B}^{-1} + \mathbf{K}^{\mathrm{T}}\mathbf{E}^{-1}\mathbf{K}\right)^{-1}\mathbf{K}^{\mathrm{T}}\mathbf{E}^{-1}\mathbf{K} \tag{8}$$

Measures of the spread of significant elements in the rows of $\mathbf{R}$ have been used to estimate the resolution of the retrieval (Conrath 1972; Menke 1984), but some authors (Thompson 1982; Huang *et al.* 1992) have reported difficulties with this approach. Purser and Huang (1993) reviewed a number of formulae to measure the spread, but found none of them to be satisfactory. Plots of spreads vs. a signal-to-noise parameter (Purser and Huang figure 5) showed significant disagreement between the different formulae, particularly in the asymptotic behaviour for small signal-to-noise values. It is difficult to devise spread formulae which give meaningful measures of resolution in the case that $\mathbf{R}$ contains significant negative elements, and there is no general agreement as to which formula to adopt.

A general problem with using any measure of the spread of rows of $\mathbf{R}$ as an estimate of the resolution, where $\mathbf{R}$ is defined by equation 8, may be highlighted by considering the case of large measurement noise*. The rows of Conrath's $\mathbf{R}$ (equation 5) are normalised so that $\sum_j [\mathbf{R}]_{ij} = 1$, but it is clear from equation 8 that row sums using this definition decrease with increasing measurement noise. Let $\mathbf{E}$ be replaced by $\epsilon\mathbf{E}'$ then, in the limit as $\epsilon \to \infty$

$$\mathbf{R} \to \epsilon^{-1}\mathbf{B}\,\mathbf{K}^{\mathrm{T}}\mathbf{E}'^{-1}\mathbf{K} \tag{9}$$

The spreads of the rows of $\mathbf{R}$, being independent of the scalar multiplier ϵ^{-1}, have well defined values which depend on $\mathbf{K}$. However, as the retrieved profile in this limit is the prior profile, with no contribution from the measurements, what do the spreads measure?

* This was drawn to my attention by Lee (1998a).

Thompson (1982) introduced an empirical resolution by considering the retrieval of two adjacent gaussian perturbations in the atmospheric temperature profile. He investigated the separability of the peaks in the retrieved profile by comparing the size of the dip between them with the retrieval error. This approach avoids the problems associated with spread functions and the resolution matrix. Statistical tests of the detectability of perturbations in the atmospheric temperature profile are discussed in section 6.

4. Degrees of freedom for signal

Purser and Huang (1993) base their analysis on the degrees of freedom for signal (DFS). This global quantity indicates the number of effectively independent pieces of information provided by the measurement. The DFS is the trace of $\mathbf{R}$ as defined by equation 8.

$$\mathrm{DFS} = \mathrm{tr}\,(\mathbf{R}) \tag{10}$$

The DFS has many desirable properties (Wahba 1985) such as independence of the units or coordinate transformations. It shares the property of information measures that the sum of the DFSs for two sets of independent measurements is never less than the individual DFSs and never more than their sum. Although the DFS is derived from the resolution matrix, the magnitudes of the elements are not normalised as in the spread formulae and the DFS correctly tends to zero in the large noise limit (equation 9).

Purser and Huang noted that the diagonal elements of $\mathbf{R}$ apportion the DFS to particular altitudes and so were able to define an effective data density over altitude. This important idea forms the basis of the analyses presented in this paper. However, the data density is not an indication of resolution; a high value might indicate an ability to measure over a wide range of resolutions with moderate precision, but could equally indicate an ability to measure only at low resolution but with high precision. In this paper I show that this ambiguity may be removed by a transformation of coordinates which allows the DFS and other global performance measures to be apportioned over both altitude and scale of structure.

5. Radiometer models

Theoretical results will be illustrated through their application to two radiometer models.

(*a*) *Ideal radiometer model.*

The upwelling radiance at the top of the atmosphere may be written as a weighted integral of the Planck function over height. A simple form for the weighting function $K_i(p)$ for radiometer channel i may be derived for the case of a monochromatic measurement in the wings of a pressure broadened line (Houghton *et al.* 1984).

$$K_i(p) = \frac{2}{h_0}\left(\frac{p}{p_i}\right)^2 \exp\left\{-\left(\frac{p}{p_i}\right)^2\right\} \tag{11}$$

This function peaks at atmospheric pressure $p = p_i$. With scale height h_0 and surface pressure p_s, the vertical coordinate h is given by $h = -h_0 \ln(p/p_s)$. The value $h_0 = 7\,\mathrm{km}$ corresponds to an atmospheric temperature of about 240 K, an average temperature for the lower 40 km, making h approximately equal to the altitude in km. The weighting function is normalised so that $\int_{-\infty}^{\infty} K\,\mathrm{d}h = 1$. At millimetre wavelengths, the Planck function

is approximately proportional to temperature and equation 1 arises on evaluating the integral over height by numerical quadrature. If the atmosphere over the altitude range 0–40 km is represented by profiles tabulated at 64 regularly spaced points so that the tabulation interval is 0.625 km, the elements of the Jacobian matrix $\mathbf{K}$ are

$$[\mathbf{K}]_{ij} = \frac{1}{5.6} \exp\left\{ - \exp\left(\frac{h_i - j}{5.6} \right) + \frac{h_i - j}{5.6} \right\} \tag{12}$$

The peak altitudes h_i for 10 functions were spread evenly over the 64 tabulation points with

$$h_i = 3.2,\ 9.6,\ 16.0,\ 22.4,\ 28.8,\ 35.2,\ 41.6,\ 48.0,\ 54.4,\ 60.8 \tag{13}$$

The peak separation is 6.4 tabulation intervals or 4 km. This distribution of the weighting functions is roughly equivalent to that for the Advanced Microwave Sounder Unit (AMSU)(Patel and Mentall 1993). No allowance was made for radiation from the Earth's surface or from the atmosphere above 40 km. Measurements were assumed to be independent with standard deviation 0.3 K so that $\mathbf{E}$ is diagonal with all diagonal elements equal to $0.09\,\mathrm{K}^2$.

(*b*) *IASI model.*

The Infrared Atmospheric Sounding Interferometer (IASI) (Diebel *et al.* 1996) covers the spectral range from $645\,\mathrm{cm}^{-1}$ (15.5 μm) to $2760\,\mathrm{cm}^{-1}$ (3.62 μm) with an approximate spectral resolution of $0.35\,\mathrm{cm}^{-1}$ unapodised and $0.5\mathrm{cm}^{-1}$ apodised. I am indebted to A. C. L. Lee of the UK Meteorological Office for supplying the matrix $\mathbf{K}^T\mathbf{E}^{-1}\mathbf{K}$ for IASI. Radiometric calculations used a fast radiative transfer code RTIASI (Matricardi and Saunders 1999) developed by the European Centre for Medium Range Forecasting (ECMWF). The Floyd atmospheric profile (appendix B) was used as a basis for evaluating the temperature derivatives constituting the Jacobian matrix $\mathbf{K}$. The Gaussian apodised Jacobians output by the RTIASI code were unapodised (Lee 1999). The noise equivalent temperature for the self-apodised spectrum was that presented by Cayla (1996) at the Cannes meeting of the IASI Sounding Working Group held in March 1997. A diagonal forward modelling error variance of $0.04\,\mathrm{K}^2$ was added to all channels.

The RTIASI code uses 43 pressure levels which are listed in appendix B. Elements in the rows of $\mathbf{K}$ were interpolated onto the 64 levels equally spaced in $\ln(p)$ used in this paper. Each element was divided by the average thickness of the adjacent layers before cubic interpolation.

6. The detectability of atmospheric perturbation patterns

Consider the probability of error in detecting the presence or absence of a particular structure in the atmospheric profile. Let $\mathbf{v}$ be a unit magnitude perturbation pattern and let α be the projection of the atmospheric state onto this pattern so that $\alpha = \mathbf{v}^{\mathrm{T}}\mathbf{x}$. The prior standard deviation σ^b of α and the post measurement standard deviation σ^a are given by

$$(\sigma^b)^2 = \mathbf{v}^{\mathrm{T}}\mathbf{B}\mathbf{v} \qquad\qquad (\sigma^a)^2 = \mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{v} \tag{14}$$

Prunet *et al.* (1998) advocated using the ratio of post measurement to prior variances as a performance indicator (their σ_a in my notation is σ^a/σ^b with their $\delta\mathbf{x}$ equal to my $\mathbf{B}\mathbf{v}$). Whilst Prunet *et al.* considered perturbation patterns derived from particular meteorological situations, Lee (1998b; 1998c) suggested the use of 'double Dirac' patterns which are identical to Haar wavelets illustrated in figure 1 and described in appendix C.

He argues that a perturbation pattern which is to be used to determine the resolution of an instrument should have zero mean value. Such a pattern is detected through its first moment and the sloping edges of the weighting functions (rows of $\mathbf{K}$). The Haar wavelet is the simplest such pattern and this is the perturbation pattern used in the remainder of this paper although similar results would be expected using other wavelets with non-zero first moments (Burrus and Gopinath 1998).

Consider the problem of deciding between two mutually exclusive hypotheses – pattern $\mathbf{x}^p$ present or pattern $\mathbf{x}^q$ present (one of these may be the null pattern). Given a measurement $\mathbf{y}$, a strategem for deciding between the hypotheses is to divide the N dimensional vector space of the measurements into two regions, $\mathcal{P}$ and $\mathcal{Q}$ and to choose $\mathbf{x}^p$ present if $\mathbf{y} \in \mathcal{P}$ and $\mathbf{x}^q$ present if $\mathbf{y} \in \mathcal{Q}$. Let the probability of deducing $\mathbf{x}^p$ present when the true pattern is $\mathbf{x}^q$ be $P_E(p|q)$. In the case that $P_E(p|q) = P_E(q|p) = P_E$, the error probability P_E is minimised when $\mathcal{P}$ includes all points where $\mathrm{N}_N(\mathbf{y} - \mathbf{K}\mathbf{x}^p, \mathbf{E}) > \mathrm{N}_N(\mathbf{y} - \mathbf{K}\mathbf{x}^q, \mathbf{E})$ and $\mathcal{Q}$ points where $\mathrm{N}_N(\mathbf{y} - \mathbf{K}\mathbf{x}^p, \mathbf{E}) \leq \mathrm{N}_N(\mathbf{y} - \mathbf{K}\mathbf{x}^q, \mathbf{E})$. The discriminator or boundary between $\mathcal{P}$ and $\mathcal{Q}$ is a hyper-plane and

$$P_E = \int_{\mathcal{Q}} \mathrm{N}_N(\mathbf{y} - \mathbf{K}\mathbf{x}^p, \mathbf{E})\, \mathrm{d}^N\mathbf{y} = \int_{\mathcal{P}} \mathrm{N}_N(\mathbf{y} - \mathbf{K}\mathbf{x}^q, \mathbf{E})\, \mathrm{d}^N\mathbf{y} \tag{15}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\Delta/2}^{\infty} e^{-t^2/2}\, \mathrm{d}t \tag{16}$$

where Δ^2 is the Mahalanobis distance between the patterns (Wald 1944; Anderson 1971) defined by

$$\Delta^2 = (\mathbf{x}^p - \mathbf{x}^q)^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} \mathbf{E}^{-1} \mathbf{K} (\mathbf{x}^p - \mathbf{x}^q) \tag{17}$$

A maximum permissible value for P_E implies a minimum permissible Mahalanobis distance (for example for $P_E \leq 0.05$, $\Delta_{\min} = 3.2897$). As an example figure 1 shows the minimum amplitude in degrees K of a profile perturbation in the form of a Haar wavelet which is to be detected with less than 5% probability of error. The minimum amplitude a is plotted as a function of the altitude h of the center of the perturbation for wavelet half-widths $w = 5$, 2.5, 1.25 and 0.625 km. Calculations were for both the ideal radiometer model (section 5 (a)) and for IASI (section 5 (b)). For the ideal radiometer model, an expression for the minimum amplitude a as a function of half-width w may be found by averaging the Mahalanobis distance Δ^2 over height with an atmosphere deep enough to avoid boundary effects. The averages were evaluated by integration over the continuous weighting functions described by equation 11 and are marked on the top and bottom axes in figure 1.

$$a'^2 = \Delta_{\min}^2 h'_\delta \left\{4 \ln\cosh(w') - \ln\cosh(2w')\right\}^{-1} \tag{18}$$

$$\approx \Delta_{\min}^2 h'_\delta \frac{1}{w'^4} \quad \text{for } w' \ll 1 \tag{19}$$

Equation 18 is in terms of the dimensionless variables $a' = a/\epsilon$, $h'_\delta = h_\delta/h_0$ and $w' = w/h_0$ where $\mathbf{E} = \epsilon^2\, \mathbf{I}$, h_δ is the altitude separation of the weighting functions and h_0 is the scale height (for figure 1 $\epsilon = 0.3\,\mathrm{K}$, $\Delta_{\min} = 3.2897$, $h_\delta = 4\,\mathrm{km}$ and $h_0 = 7\,\mathrm{km}$). For small values of w, the instrument responds to the first moment aw^2 of the perturbation (equation 19 gives values of a accurate to 10% for $w' < 0.41$). Equation 18 shows how the minimum amplitude for detection depends on properties of the instrument and of radiative transfer in the atmosphere. However this equation is based on an idealised model with equally spaced, identically shaped weighting functions. Many absorption lines contribute to the absorption at most wavelengths and the assumption that the absorption

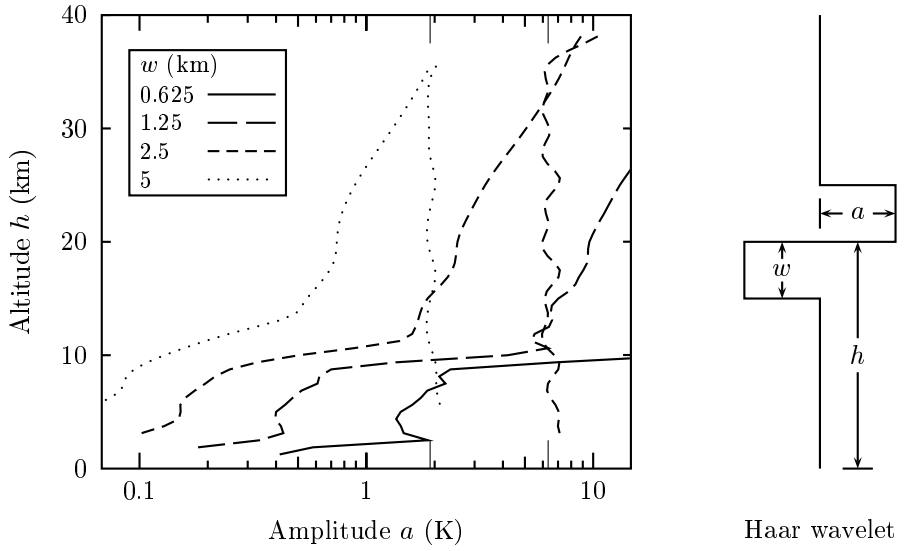


Figure 1. *The Haar wavelet and minimum amplitudes for at least 95% certainty of detection.* Amplitudes are plotted as a function of height for IASI (sloping lines) and for the ideal radiometer model (vertical lines). Average values for the ideal radiometer calculated using equation 18 are marked on the top and bottom axes.

coefficient varies as the square of atmospheric pressure implicit in equation 12 is true only in the wings of a single isolated line. The simple model may be improved to some extent by using an absorption coefficient proportional to p^β with $\beta < 2$. Equation 11 becomes $K_i(p) = (\beta/h_0)(p/p_i)^\beta \exp[-(p/p_i)^\beta]$ resulting in a wider weighting function. Equation 18 is then modified by substituting $2h_0/\beta$ for h_0 in the definitions of h'_δ and w'. Although equations 18 and 19 give insight into how the minimum detectable amplitude depends on model parameters, generally it is not practical to estimate these parameters for a radiometer such as IASI.

The minimum detectable amplitudes for IASI are about a factor 40 smaller than those for the ideal radiometer in the troposphere, but increase through the stratosphere to reach comparable values by 40 km altitude. This reflects the high density of IASI channels with weighting functions peaking in the troposphere. Up to 7.5 km altitude the minimum amplitude a for a wavelet with half-width $w = 1$ km to be detected with at least 95% certainty is 0.9 K. Prunet *et al.* (1998) refering to IASI conclude that 'fine scale structures with a vertical resolution of 1 km are estimated with standard deviation 0.7 K'. For a cumulative normal distribution, the corresponding amplitude for detection with 95% certainty is 1.2 K. Since Prunet *et al.*'s 'fine scale structure' is not a Haar wavelet, this agreement is surprisingly close.

The requirement for a minimum probability of error in detecting the presence of a pattern in the atmospheric profile may be generalised to the detection of any standardised linear combination of a set of L patterns forming the columns of an $M \times L$ matrix $\mathbf{X}$. For any L-dimensional unit vector $\mathbf{u}$

$$\mathbf{u}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{X}\mathbf{u} \geq \Delta^2_{\min} \tag{20}$$

Let $\mathbf{\Lambda}$ be the diagonal matrix of the eigenvalues of $\mathbf{X}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{X}$ and the columns of $\mathbf{U}$ its

eigenvectors*. Then

$$\mathbf{u}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{X}\mathbf{u} = \mathbf{u}^{\mathrm{T}}\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}}\mathbf{u} = \mathbf{u'}^{\mathrm{T}}\mathbf{\Lambda}\mathbf{u'} = \sum_{i=1}^{L} [\mathbf{u'}]_i^2 \, \lambda_i \tag{21}$$

where $\mathbf{u'} = \mathbf{U}^{\mathrm{T}}\mathbf{u}$ is another L-dimensional unit vector and λ_i are the eigenvalues of $\mathbf{X}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{X}$. Inequality 20 is satisfied for every unit vector $\mathbf{u'}$ if and only if the minimum eigenvalue $\lambda_{\min}$ satisfies

$$\lambda_{\min} \geq \Delta_{\min}^2 \tag{22}$$

The problem of detecting one of many possible patterns which may be simultaneously present does not have a simple solution comparable with equations 16 and 17. However, in the case that the patterns are closely spaced, the probability of a correct choice is maximised by maximising the probability densities at their mean values; that is by minimising $|\mathbf{A}|$, the determinant of the post-measurement covariance matrix. This is equivalent to maximising the entropic information H (section 8).

7. Multiresolution analysis

It was noted in section 4 that global performance measures such as the DFS are unaffected by orthogonal coordinate transformations. For instance a complete orthonormal M dimensional set is formed by vectors whose ith components are $\sqrt{2/M}\cos(2\pi ni/M)$ and $\sqrt{2/M}\sin(2\pi ni/M)$ for $n = 1,\ 2,\ \ldots,\ (M/2) - 1$, together with vectors with components $\sqrt{1/M}$ and $\sqrt{1/M}(-1)^i$. If these vectors are used as the columns of an orthogonal matrix $\mathbf{V}$, the coordinate transform $\mathbf{w} = \mathbf{V}^{\mathrm{T}}\mathbf{x}$ constitutes a spatial Fourier transform of the atmospheric state $\mathbf{x}$, each component of $\mathbf{w}$ giving the amplitude of a Fourier component. Recalling that the diagonal elements of the resolution matrix $\mathbf{R}$ apportioned the DFS over altitude (section 4), it is clear that the diagonal elements of the transformed resolution matrix $\mathbf{V}^{\mathrm{T}}\mathbf{R}\mathbf{V}$ apportion the DFS over spatial frequencies. The global DFS is unchanged by the transformation since $\mathrm{DFS} = \mathrm{tr}(\mathbf{R}) = \mathrm{tr}(\mathbf{V}^{\mathrm{T}}\mathbf{R}\mathbf{V})$. The Fourier transform gives information about resolution (spatial frequencies), but all information about altitude is lost.

Wavelets have both position (altitude) and scale. By transforming to a wavelet basis, it is possible to apportion global performance measures such as the DFS simultaneously over both altitude and scale giving information about the resolution as a function of altitude.

Consider a set of unit vectors which are scaled versions of the Haar wavelet shown in figure 1. The combination of wavelet altitude h and width w for each vector may be chosen so that a set of $M - 1$ such vectors, together with a vector each component of which is equal to $\sqrt{1/M}$, form a complete orthonormal set. Details of these Haar wavelet vectors are described in appendix C, but the following analysis is not specific to the Haar wavelet; many other types have been proposed (Burrus and Gopinath 1998). The wavelet vectors form columns of an orthogonal transformation matrix $\mathbf{V}$. The vector $\mathbf{w}$ of wavelet amplitudes where

$$\mathbf{w} = \mathbf{V}^{\mathrm{T}}\mathbf{x} \tag{23}$$

constitutes a multiresolution analysis of the atmospheric state. Each component of $\mathbf{w}$ quantifies the contribution to $\mathbf{x}$ of the pattern described by the corresponding wavelet which is characterised by particular values of altitude h and width w.

* Prior information has been included here so that $\mathbf{A}^{-1}$ replaces $\mathbf{K}^{\mathrm{T}}\mathbf{E}^{-1}\mathbf{K}$ in equation 17.

This transformation could be applied to $\mathbf{R}$ to give a multiscalar analysis of the DFS. However, in the next section it will be shown that Shannon's entropic information gain H is an appropriate global performance measure for remote sounding radiometers. Multiresolution analysis will be applied to H in section 9.

8. Information measures

The Kullback-Leibler entropy G (Kullback 1978) is a measure of the information in favour of an hypothesis $\mathcal{H}_1$ against $\mathcal{H}_2$. If the probability distribution of the measurements under hypothesis $\mathcal{H}_i$ is $P_i(\mathbf{y})$

$$G = \int P_1(\mathbf{y}) \log \frac{P_1(\mathbf{y})}{P_2(\mathbf{y})} \, \mathrm{d}^N\mathbf{y} \tag{24}$$

The logarithms in information expressions may have base 2 to give information in bits or base e to give information in natural units*. Let $\mathcal{H}_1$ correspond to the case that the atmospheric profile is perturbed by an amount $\mathbf{x}$ so that $P_1(\mathbf{y}) = P_e(\mathbf{y} - \mathbf{K}\mathbf{x})$ where $P_e(\mathbf{y}) = \mathrm{N}_N(\mathbf{y}, \mathbf{E})$, the distribution of measurement errors. Consider the following two cases for $\mathcal{H}_2$:

Case A: The atmospheric profile is unperturbed so that $P_2(\mathbf{y}) = P_e(\mathbf{y})$. Then from equation 24

$$G = \frac{\log(e)}{2} \mathbf{x}^{\mathrm{T}} \mathbf{K}^{\mathrm{T}} \mathbf{E}^{-1} \mathbf{K}\mathbf{x} = \frac{\log(e)}{2} \Delta^2 \tag{25}$$

where Δ^2 is the Mahalanobis distance described in section 6. The average information $\langle G \rangle_{\mathbf{x}}$ over perturbations distributed as $P_b(\mathbf{x}) = \mathrm{N}_M(\mathbf{x}, \mathbf{B})$ is

$$\langle G \rangle_{\mathbf{x}} = \int P_b(\mathbf{x}) G \, \mathrm{d}^M\mathbf{x} = \frac{\log(e)}{2} \mathrm{tr} \left(\mathbf{B}\mathbf{K}^{\mathrm{T}} \mathbf{E}^{-1} \mathbf{K} \right) \tag{26}$$

Here $\mathbf{K}^{\mathrm{T}} \mathbf{E}^{-1} \mathbf{K}$ is the Fisher information matrix† associated with the measurement of the atmospheric profile.

Case B: The atmospheric profiles are randomly perturbed with distribution $P_b(\mathbf{x}) = \mathrm{N}_M(\mathbf{x}, \mathbf{B})$ so that $P_2(\mathbf{y}) = \int P_b(\mathbf{x}) P_e(\mathbf{y} - \mathbf{K}\mathbf{x}) \, \mathrm{d}^M\mathbf{x} = \mathrm{N}_N(\mathbf{y}, \mathbf{E} + \mathbf{K}\mathbf{B}\mathbf{K}^{\mathrm{T}})$. The average Kullback-Leibler information is then

$$\begin{aligned}
\langle G \rangle_{\mathbf{x}} &= \int P_b(\mathbf{x}) \, G \, \mathrm{d}^M\mathbf{x} \\
&= \iint P_b(\mathbf{x}) P_e(\mathbf{y} - \mathbf{K}\mathbf{x}) \log P_e(\mathbf{y} - \mathbf{K}\mathbf{x}) \, \mathrm{d}^M\mathbf{x} \, \mathrm{d}^N\mathbf{y} \\
&\quad - \iint P_b(\mathbf{x}) P_e(\mathbf{y} - \mathbf{K}\mathbf{x}) \log P_2(\mathbf{y}) \, \mathrm{d}^M\mathbf{x} \, \mathrm{d}^N\mathbf{y} \\
&= \int P_e(\mathbf{y}) \log P_e(\mathbf{y}) \, \mathrm{d}^N\mathbf{y} - \int P_2(\mathbf{y}) \log P_2(\mathbf{y}) \, \mathrm{d}^N\mathbf{y} \\
&= H
\end{aligned} \tag{27}$$

Here H is the Shannon entropic information gain for the measurement (Shannon 1949; Lindley 1956). For the normal probability distributions assumed here $H = \frac{1}{2} \log \frac{|\mathbf{E} + \mathbf{K}\mathbf{B}\mathbf{K}^{\mathrm{T}}|}{|\mathbf{E}|}$

* Logarithms in information expressions which may have base 2 or e will be written as 'log'. Elsewhere logarithms have base e and are written 'ln'.

† The (i, j) th element of the Fisher information matrix (Fisher 1956) is $\int \frac{1}{P_1} \frac{\partial P_1}{\partial x_i} \frac{\partial P_1}{\partial x_j} \, \mathrm{d}^N\mathbf{y}$.

which may be written $H = \frac{1}{2}\log\frac{|\mathbf{B}|}{|\mathbf{A}|}$ where $\mathbf{B}$ is the prior atmospheric covariance matrix and $\mathbf{A}$ is the covariance matrix after the measurement[‡].

An analysis of radiometer performance using Fisher information (case A above) would be similar to the analysis already undertaken in section 6. In this case the instrument is being used to distinguish between only two possible atmospheric profiles. Case B applies to the more realistic requirement to detect the presence of one particular perturbation profile in a more generally perturbed atmosphere.

A simple argument relates Shannon's information gain to probabilities for the multivariate normal distributions considered here. Suppose that the profiles to be detected form a closely spaced rectangular lattice in the M-dimensional vector space of $\mathbf{x}$. If the volume of a cell of this lattice is $\delta^M\mathbf{x}$, the post measurement probability of correctly identifying a profile is P^a where

$$\lim_{\delta^M\mathbf{x}\to 0} P^a = \delta^M\mathbf{x}\,\mathrm{N}_M(\mathbf{0},\mathbf{A}) \propto \delta^M\mathbf{x}\,|\mathbf{A}|^{-1/2} \tag{28}$$

The ratio of the post measurement P^a to the prior P^b probability of correct identification remains finite in the limit and the logarithm of this ratio is equal to Shannon's information gain.

$$\lim_{\delta^M\mathbf{x}\to 0} \log\frac{P^a}{P^b} = \frac{1}{2}\log\frac{|\mathbf{B}|}{|\mathbf{A}|} = H \tag{29}$$

A relationship between Shannon's information and the variance of the principal components is derived in appendix D.

Shannon's information gain was introduced by Peckham (1974) as a measure of the quality of remote sensing measurements and was used by Eyre (1990) in a simulation study of satellite sounding systems.

9. Multiresolution information

Multiresolution analysis may be applied to the resolution matrix (section 4), to Fisher information or to Shannon's entropic information (section 8). Information, like the DFS, may be expressed as the trace of a matrix so distributing information over the coordinates. Huang and Purser (1996) used the fact that the logarithm of the determinant of a matrix is equal to the trace of the matrix logarithm* to express the information gain H as the trace of an information gain matrix. This allowed them to define an information density over altitude.

$$H = \mathrm{tr}(\mathbf{H}^{\mathrm{HP}}) \tag{30}$$

where

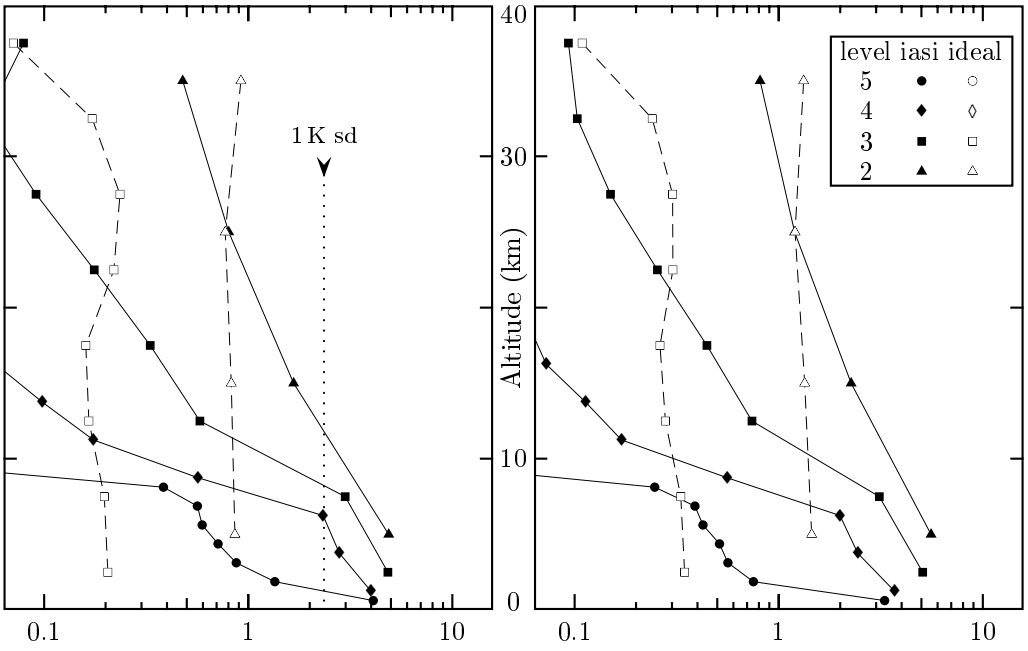
$$\mathbf{H}^{\mathrm{HP}} = \frac{1}{2}\log(\mathbf{B}\mathbf{A}^{-1}) \tag{31}$$

The diagonal element $[\mathbf{H}^{\mathrm{HP}}]_{ii}$ is the information gain attributable to the altitude corresponding to the ith component of the atmospheric state vector. It is preferable to work with the matrix $\mathbf{H}$ defined by

$$\mathbf{H} = \frac{1}{2}\{\log(\mathbf{B}) - \log(\mathbf{A})\} \tag{32}$$

[‡] Substituting for $\mathbf{A}$ from equation 3, the two forms for H are equivalent if $|\mathbf{I}+\mathbf{B}\mathbf{K}^T\mathbf{E}^{-1}\mathbf{K}| = |\mathbf{I}+\mathbf{E}^{-1}\mathbf{K}\mathbf{B}\mathbf{K}^T|$. This is a special case of the determinant identity $|\mathbf{I}+\mathbf{X}\mathbf{Y}| = |\mathbf{I}+\mathbf{Y}\mathbf{X}|$ where $\mathbf{X}$ is $M\times N$ and $\mathbf{Y}$ is $N\times M$ (see, for example, Gantmacher 1959).

* The logarithm of a square matrix may be found by replacing the eigenvalues by their logarithms in an eigenvalue expansion of the matrix (Golub and Van Loan 1996).



(a) *Information gain: diagonal prior* (b) *Information gain: non-diagonal prior*

Figure 2. *Wavelet components of the multiresolution information gain.* The vertical scale covers the altitude range from 0–40 km. The horizontal scales are information gain in binary units (bits). The left-hand graph (a) is based on a diagonal prior covariance matrix. The right hand graph (b) is for the non-diagonal prior covariance described by equation 35. The information gain for both the simple model and for IASI described in section 5 is plotted. The vertical dotted line in graph (a) shows the information gain for a simplified measurement in which all components are measured with 1 K standard deviation noise.

as, although both forms preserve the additivity of total information, only $\mathbf{H}$ preserves the additivity of information separately for each diagonal element. For non-commuting matrices $\mathbf{H} \neq \mathbf{H}^{\mathrm{HP}}$ although $\mathrm{tr}(\mathbf{H}) = \mathrm{tr}(\mathbf{H}^{\mathrm{HP}})$. Huang and Purser distributed information gain over altitude but not over scales of patterns in the atmospheric profile. A multiresolution wavelet transform allows information to be distributed over both altitude and scale. The covariances transform to $\mathbf{V}^{\mathrm{T}}\mathbf{A}\mathbf{V}$ and $\mathbf{V}^{\mathrm{T}}\mathbf{B}\mathbf{V}$. The total information gain H, which is unchanged by this transformation, may be expressed as the trace of a multiresolution information gain matrix.

$$H = \mathrm{tr}\left(\mathbf{H}^{\mathrm{W}}\right) \tag{33}$$

where

$$\mathbf{H}^{\mathrm{W}} = \frac{1}{2}\left\{\log(\mathbf{V}^{\mathrm{T}}\mathbf{B}\mathbf{V}) - \log(\mathbf{V}^{\mathrm{T}}\mathbf{A}\mathbf{V})\right\} = \frac{1}{2}\mathbf{V}^{\mathrm{T}}\left\{\log(\mathbf{B}) - \log(\mathbf{A})\right\}\mathbf{V} \tag{34}$$

The diagonal element $[\mathbf{H}^{\mathrm{W}}]_{ii}$ represents the information gain at the altitude and scale of the wavelet forming the ith column of $\mathbf{V}$.

Equation 34 was used to calculate the multiresolution information gain for the two models described in section 5. The diagonal elements $[\mathbf{H}^{\mathrm{W}}]_{ii}$ are plotted in figure 2, as described in the caption, for wavelet levels $j = 2$–5 corresponding to half widths $w = 5$–0.625 km ($w = 20 \times 2^{-j}$ km). The vertical axis represents altitude over the range 0–40 km and the point representing each wavelet component is plotted at the altitude of the centre of the wavelet. The left hand graph (figure 2 (a)) is for a diagonal prior covariance

$\mathbf{B} = 25\,\mathbf{I}\,\mathrm{K}^2$. The information gain per component corresponding to a hypothetical measurement by which all components of the profile are determined to a standard deviation precision of 1 K is 2.35 bits. For comparison this value is shown by a vertical dotted line.

The right hand graph is for the prior covariance used by Huang and Purser (1996). They assumed an exponential prior covariance with a width varying linearly from 3 km at 8 km altitude to 8 km at 30 km (values read from their figure 2). With a uniform standard deviation of 5 K at all altitudes, this is approximated in units of K^2 by

$$[\mathbf{B}]_{ij} = 25 \exp\left(-\frac{|i-j|}{2 + 0.1(i+j)}\right) \tag{35}$$

The tropospheric information gain for IASI is very much greater than that for the ideal radiometer model described in section 5. This is to be expected as the ideal model includes only 10 measurements, and, without prior information, these would allow only 10 wavelet components of the atmospheric profile to be determined. Wavelet levels 0–2, together with the profile average, include 8 values. Assuming that these levels are better determined by the measurements than those requiring higher resolution, there is very little information left from the measurements for levels 3–5. The effect of correlation in the prior covariance matrix is shown by the increase in the information gain in figure 2 (b) compared with 2 (a). (The prior covariance matrices for (a) and (b) have equal diagonal elements, but that for (b) described by equation 35 includes correlation elements.)

10. Conclusions

The performance of atmospheric sounding radiometers may be described through their ability to reliably detect perturbations of the atmospheric temperature profile in the form of Haar wavelets. Haar functions have zero mean value and may have any desired half-width and altitude so that they are ideally suited to analysing the resolution of a radiometer as a function of altitude. Two scenarios have been considered:

1. The perturbation is to be detected against an unperturbed atmosphere. In this case the probability of detection depends on the Mahalanobis distance.
2. The specific perturbation is to be detected against a randomly perturbed atmosphere. In this case the average information gain is given by Shannon's entropic information. A multiresolution wavelet analysis allows the total information gain to be distributed over altitude and resolution.

IASI has been shown to have a 95% probability of detecting perturbations in the troposphere up to 7.5 km altitude with half width 1 km and amplitude 0.9 K. An ideal model radiometer, similar in performance to the AMSU, required a perturbation with half width 2.5 km and amplitude 6 K to achieve the same probability of detection.

Appendix A

Notation

Recently there has been an attempt to introduce a unified notation in atmospheric and oceanic data assimilation (Ide *et al.* 1997). I have tried to follow this as far as practicable, but conflicts arise with other well established notations, notably $\mathbf{R}$ for the resolution matrix and H for entropy and entropic information. Major notation used in this paper is summarised in table 1.

TABLE 1. NOTATION

Symbol	Dimension	Description
$\mathbf{A}$	$M \times M$	Post measurement covariance of $\mathbf{x}$
$\mathbf{B}$	$M \times M$	Covariance of prior estimate
$\mathbf{e}$	N	Measurement + model error vector
$\mathbf{E}$	$N \times N$	Covariance of $\mathbf{e}$
$\mathbf{K}$	$N \times M$	Linearised forward model coefficients (Jacobian matrix)
$\mathrm{N}_M(\cdots)$	1	Normal probability distribution (equation 4)
$\mathbf{x}$	M	Atmospheric state vector
$\mathbf{x}^{\mathrm{a}}$	M	Post-measurement (analysis) estimate of $\mathbf{x}$
$\mathbf{x}^{\mathrm{b}}$	M	Prior (base) estimate of $\mathbf{x}$
$\mathbf{y}$	N	Vector of radiance measurements
G	1	Kullback-Leibler information
H	1	Shannon's entropic information gain
$\mathbf{H}$	$M \times M$	Information gain matrix
$\mathbf{H}^{\mathrm{W}}$	$M \times M$	Multiresolution information gain matrix
$\mathbf{R}$	$M \times M$	Resolution matrix
$\mathbf{V}$	$M \times M$	Orthogonal wavelet transformation matrix
Δ^2	1	Mahalanobis distance

APPENDIX B

Floyd profile

Radiative transfer calculations for the IASI instrument (Diebel *et al.* 1996) were based on the fast radiative transfer model RTIASI developed by the European Centre for Medium Range Forecasting (Matricardi and Saunders 1999). The 43 pressure levels of this model are listed in table 2. The calculation of the temperature derivatives which form the Jacobian matrix $\mathbf{K}$ were based on the 'Floyd' profile defined in this table. This profile is that illustrated in Prunet *et al.*'s (1998) figure 2 for latitude 54 N. The matrix $\mathbf{K}^{\mathrm{T}}\mathbf{E}^{-1}\mathbf{K}$ was interpolated onto the tabulation altitudes 1–64 as described in section 5 (b).

APPENDIX C

Haar wavelets

Haar wavelets were introduced by Haar (1910) and are described in many text books such as Burrus and Gopinath (1998). Define the components of the M-dimensional Haar source vector $\mathbf{v}$ as

$$\mathbf{v}_i = \begin{cases} a & \text{for } 0 \le i < M/2 \\ -a & \text{for } M/2 \le i < M \\ 0 & \text{otherwise} \end{cases} \tag{C.1}$$

with $M = 2^L$, an integer power of two, and $a = 1/\sqrt{M}$ so that $\sum_{i=0}^{M-1} |\mathbf{v}_i|^2 = 1$. (In this appendix subscripts indicate vector components numbered from 0.) Define scaled and translated wavelets as

$$\mathbf{v}_i^{(j,k)} = 2^{j/2}\mathbf{v}_{(2^j i - Mk)} \quad \text{for } i = 0, 1, \ \ldots \ , M-1 \tag{C.2}$$

where j is the level (scale) index, k the translation index and the factor $2^{j/2}$ maintains normalisation. $\mathbf{v}^{(j,k)}$ is non-zero only over a altitude range from $kh^{(j)}$ to $(k+1)h^{(j)}$ where $h^{(j)} = 2^{L-j}h$ and h is the tabulation interval.

TABLE 2. Floyd profile

	pressure (hPa)	altitude (tabulation intervals)	temperature (K)	water vapour (ppm)	ozone (ppm)
1	0.10	103.30	230.93	4.26	0.68
2	0.29	91.38	249.66	4.27	1.44
3	0.69	81.67	265.16	4.27	2.38
4	1.42	73.59	259.12	4.16	3.71
5	2.61	66.77	252.38	3.95	5.20
6	4.41	60.90	242.04	3.67	5.91
7	6.95	55.80	234.14	3.42	5.92
8	10.37	51.32	230.02	3.07	5.63
9	14.81	47.33	227.72	3.75	5.21
10	20.40	43.74	225.78	4.12	4.62
11	27.26	40.49	223.96	4.00	3.89
12	35.51	37.53	222.11	4.02	3.12
13	45.29	34.81	220.18	4.02	2.43
14	56.73	32.29	220.10	4.02	1.90
15	69.97	29.94	221.21	4.02	1.46
16	85.18	27.73	221.54	4.02	1.01
17	102.05	25.71	221.09	4.05	0.70
18	122.04	23.71	220.65	3.80	0.46
19	143.84	21.86	221.78	4.22	0.31
20	167.95	20.13	223.13	3.91	0.22
21	194.36	18.49	223.09	3.33	0.17
22	222.94	16.96	223.39	9.44	0.13
23	253.71	15.51	225.00	12.27	0.11
24	286.60	14.14	227.50	5.73	0.09
25	321.50	12.86	230.03	35.74	0.09
26	358.28	11.64	234.58	162.29	0.08
27	396.81	10.50	237.96	158.48	0.07
28	436.95	9.42	240.30	348.90	0.07
29	478.54	8.40	245.23	444.78	0.06
30	521.46	7.44	249.67	656.51	0.06
31	565.54	6.53	253.82	880.67	0.06
32	610.60	5.67	258.11	1141.42	0.06
33	656.43	4.86	262.28	1265.48	0.05
34	702.73	4.10	265.80	1079.04	0.05
35	749.12	3.38	268.52	1068.50	0.05
36	795.09	2.72	270.80	1467.77	0.05
37	839.95	2.10	272.98	2666.07	0.05
38	882.80	1.54	274.49	5494.70	0.05
39	922.46	1.05	277.00	7616.39	0.05
40	957.44	0.63	279.98	7727.12	0.05
41	985.88	0.31	282.05	7803.32	0.05
42	1005.43	0.09	283.33	7802.74	0.05
43	1013.25	0.00	283.83	7759.24	0.05

Let $\mathbf{v}_i^{(0)} = a$ for $i = 0, 1, \ldots, M-1$. The set of M vectors

$$\mathbf{v}^{(0)}, \mathbf{v}^{(0,0)}, \mathbf{v}^{(1,0)}, \mathbf{v}^{(1,1)}, \mathbf{v}^{(2,0)}, \mathbf{v}^{(2,1)}, \mathbf{v}^{(2,2)}, \mathbf{v}^{(2,3)}, \mathbf{v}^{(3,0)}, \ldots \\ \ldots, \mathbf{v}^{(L-1,0)}, \mathbf{v}^{(L-1,1)}, \ldots, \mathbf{v}^{(L-1,M/2-1)} \tag{C.3}$$

is a complete orthonormal set which may form the columns of an orthogonal matrix $\mathbf{V}$. (Note that at level j, there are 2^j translated wavelets so that $\mathbf{V}$ has $1 + 1 + 2 + 2^2 + \cdots + 2^{L-1} = 2^L = M$ columns.) The unnormalised Haar wavelet vectors of dimension $M = 8$ are shown in table 3. The extension to higher dimensions is obvious.

TABLE 3. HAAR WAVELETS OF DIMENSION 8 (UNNORMALISED)

Level 0		Level 1		Level 2			
1	1	1	0	1	0	0	0
1	1	1	0	−1	0	0	0
1	1	−1	0	0	1	0	0
1	1	−1	0	0	−1	0	0
1	−1	0	1	0	0	1	0
1	−1	0	1	0	0	−1	0
1	−1	0	−1	0	0	0	1
1	−1	0	−1	0	0	0	−1

APPENDIX D

Entropic information and principal component analysis

Apart from a constant, the entropic information H^a associated with a multivariate normal distribution $\mathrm{N}_M(\mathbf{x}-\mathbf{x}^{\mathrm{a}}, \mathbf{A})$ may be written as a sum over the eigenvalues λ_k of $\mathbf{A}$

$$H^a = -\frac{1}{2}\log|\mathbf{A}| = -\frac{1}{2}\sum_{k=1}^{M}\log\lambda_k \tag{D.1}$$

If $\mathbf{u}_k$ is the eigenvector* and $\mathbf{v}_i$ is the wavelet forming the i th column of the orthogonal wavelet transform matrix $\mathbf{V}$, equation 34 implies

$$H^a = -\frac{1}{2}\sum_{i=1}^{M}\sum_{k=1}^{M}(\mathbf{v}_i\cdot\mathbf{u}_k)^2\log\lambda_k \tag{D.2}$$

The information associated with each eigenvalue contributes an amount to the i th component of a multiresolution information analysis proportional to the square of the projection of the corresponding eigenvector onto the i th wavelet.

A link may be made between Prunet *et al.*'s (1998) principal component analysis of their matrix $\mathbf{A}' = \mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$ and the entropic information gain H. Writing H as a sum over the eigenvalues λ_k of $\mathbf{A}'$

$$H = -\frac{1}{2}\mathrm{tr}\log\left(\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}\right) \tag{D.3}$$

$$= \frac{1}{2}\sum_{i,k}[\mathbf{U}]_{ik}^2(-\log\lambda_k) \tag{D.4}$$

Here $[\mathbf{U}]_{ik}$, $i=1, 2, \ldots, M$ are the components of the k th eigenvector of $\mathbf{A}'$. Since $\sum_i[\mathbf{U}]_{ik}^2 = 1$, the coefficient $[\mathbf{U}]_{ik}^2$ is that fraction of the information associated with the k th eigenvector which is attributable to the altitude corresponding to index i. The maximum value for λ_k is 1 indicating no information gain. Prunet *et al.* discard eigenvectors with eigenvalues greater than a threshold of 0.9. Define $c_k = 0$ for $\lambda_k > 0.9$ and $c_k = 1$ otherwise, then the components of Prunet *et al.*'s 'integrated eigenvector' IEV are

$$\mathrm{IEV}_i = \sum_k c_k[\mathbf{U}]_{ik}^2 \tag{D.5}$$

These components satisfy $\sum_i \mathrm{IEV}_i = K$, where K is the number of retained eigenvectors. The i th component is referred to as the 'normalised resolution' at altitude i in Prunet

* The principal components are the independent random variables $\mathbf{u}_k\cdot\mathbf{x}$ with variances λ_k.

et al.'s figures 6 and 9. The usefulness of the IEV as a measure of resolution depends on the assumption that the discarded eigenvectors ($\lambda_k > 0.9$) have a finer vertical structure than those retained. In general this appears to be true as, although figure 5 of Prunet *et al.* referring to the IASI instrument does not show discarded eigenvectors, it does show a systematic decrease in the scale of the structure with increasing eigenvalues. To interpret the IEV in absolute terms, it is necessary to examine the scale of the finest structure in the retained eigenvectors.

Acknowledgements

My interest in the topic of this paper was stimulated by discussions with staff of the U.K. Meteorological Office (particularly with A. C. L. Lee) in connection with the interpretation of data from their airborne interferometric radiometer. I am indebted to Lee for supplying the matrix $\mathbf{KE}^{-1}\mathbf{K}$ for IASI. I also thank the editor, J.-N. Thépaut, and referees P. Prunet and anonymous, for their helpful suggestions. Numerical computations were carried out using the *Scilab* scientific software package (Gomez 1999).

References

Anderson, T. W.	1971	*An introduction to multivariate statistical analysis.* John Wiley, New York
Aumann, H. H. and Pagano, R. J.	1994	Atmospheric Infrared Sounder on the Earth Observing System. *Opt. Eng.*, **33,** 776–784
Backus, G. E. and Gilbert, J. F.	1970	Uniqueness in the inversion of inaccurate gross earth data. *Phil. Trans. Roy. Soc.*, **A266,** 123–192
Burrus, S. C. and Gopinath, R A.	1998	*Wavelets and wavelet transforms.* Prentice Hall, New Jersey, USA
Cayla, F.	1996	'IASI Level 1 Data'. Technical Report IA-TN-0000-5479-CNE, CNES, France
Conrath, B. J.	1972	Vertical resolution of temperature profiles obtained from remote radiation measurements. *J. Atmos. Sci.*, **29,** 1262–1271
Diebel, D., Cayla, F. and Phulpin, T.	1996	'Mission rationale and requirements'. CNES/EUMETSAT Report IA-SM-0000-10-CNE/EUM, Issue 3
Eyre, J. R.	1990	The information content of data from satellite sounding systems: A simulation study. *Q. J. R. Meteorol. Soc.*, **116,** 401–434
Fisher, R. A.	1956	*Statistical methods and scientific inference.* Oliver and Boyd, London
Gantmacher, F. R.	1959	*The Theory of Matrices. Volume 1.* Chelsea, New York
Golub, G. H. and Loan, C. F. Van	1996	*Matrix Computations.* Johns Hopkins University Press, Baltimore, USA
Gomez, C.	1999	*Engineering and Scientific Computing with Scilab.* Birkhauser, Boston
Haar, A.	1910	Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, **69**, 331–371
Houghton, J. T. and Taylor, F. W.	1973	Remote sounding from artificial satellites and space probes of the atmospheres of the earth and planets. *Rep. Prog. Phys.*, **36,** 827–919
Houghton, J. T., Taylor, F. W. and Rodgers, C. D.	1984	*Remote sounding of atmospheres.* Cambridge University Press, Cambridge
Huang, H. L., Smith, W. L. and Woolf, H. M.	1992	Vertical resolution and accuracy of infrared sounding spectrometers. *J. Appl. Meteor.*, **31,** 265–274
Huang, H. L. and Purser, R. J.	1996	Objective measures of the information density of satellite data. *Meteorol. Atmos. Phys.*, **60,** 105–117

Ide, K., Courtier, P., Ghil,M. and Lorenc, A. C.	1997	Unified notation for data assimilation: operational, sequential and variational. *Journal of the Meteorological Society of Japan*, **75,** 181–189
Kaplan, L. D.	1959	Inference of atmospheric structure from remote radiation measurements. *J. Opt. Soc. Amer.*, **49,** 1004–1007
Kullback, S.	1978	*Information theory and statistics*. Dover, Gloucester, Mass.
Lee, A. C. L.	1998a	'The resolution matrix'. Meteorological Office (Remote Sensing Instrumentation) Working Paper no. 135, held in the UK National Meteorological Library, and also obtainable through the author
Lee, A. C. L.	1998b	'Observability of atmospheric perturbation-structures by an IASI-like instrument'. Meteorological Office (Remote Sensing Instrumentation) Working Paper no. 117, held in the UK National Meteorological Library, and also obtainable through the author
Lee, A. C. L.	1998c	'Observing double-Dirac 1 km 1 K with IASI'. Meteorological Office (Remote Sensing Instrumentation) Working Paper no. 141, held in the UK National Meteorological Library, and also obtainable through the author
Lee, A. C. L.	1999	'IASI, HIRS and AGIRS'. Meteorological Office (Remote Sensing Instrumentation) Working Paper no. 160, held in the UK National Meteorological Library, and also obtainable through the author
Lindley, D. V.	1956	On a measure of the information provided by an experiment. *Ann. Mathematical Statistics*, **27,** 986–1005
Matricardi, M. and Saunders, R.	1999	Fast radiative transfer model for simulation of infrared atmospheric sounding interferometer radiances. *Appl. Opt.*, **38,** 5679–5691
Menke, W.	1984	*Geophysical data analysis: discreet inverse theory*. Academic Press, New York
Patel, P. K. and Mentall, J.	1993	'The advanced microwave sounding unit A (AMSU-A)' in 'Microwave Instrumentation for Remote Sensing of the Earth. Orlando, FL, USA. SPIE. 13-14 April 1993'. *Proceedings of SPIE*, **1935,** 130–135
Peckham, G.	1974	Remote measurements of atmospheric temperature by satellite infra-red radiometry and optimum radiometer configurations. *Quart. J. R. Meteorol. Soc.*, **100,** 406–419
Prunet, P., Thépaut, J-N. and Cassé, V.	1998	The information content of clear sky IASI radiances and their potential for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **124,** 211–241
Purser, R. J. and Huang, H. L.	1993	Estimating the effective data density in a satellite retrieval or an objective analysis. *J. Appl. Meteor.*, **32,** 1092–1107
Rodgers, C. D.	1976	Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Rev. Geophys. and Space Phys.*, **14,** 609–624
Rodgers, C. D.	2000	*Inverse methods for atmospheric sounding: theory and practise*. World Scientific Publishing
Shannon, C. E.	1949	Communication in the presence of noise. *Proc. I. C. E.*, **37,** 10–21
Simeoni, D., Singer, C. and Chalon, G.	1997	Infrared Atmospheric Sounding Interferometer. *Acta Astronautica*, **40,** 113–118

Smith, W. L., Revercomb, H. E., Laporte, D. D., Sromovsky, L. A., Silverman, S., Woolf, H. M., Howell, H. B., Knuteson, R. O. and Huang, H. L.	1990	GHIS – The GOES High Resolution Interferometer Sounder. *J. Appl. Meteor.*, **29,** 1189–1204
Thompson, O. E.	1982	HIRS-AMST satellite sounding system – Theoretical and empirical resolving power. *J. Appl. Meteor.*, **21,** 1550–1561
Wahba, G.	1985	Design criteria and eigensequence plots for satellite computed tomography. *J. Atmos. Oceanic Technol.*, **2,** 125–132
Wald, A.	1944	On a statistical problem arising in the classification of an individual into one of two groups. *Annals of Mathematical Statistics*, **15**, 145–162
Wark, D. Q. and Fleming, H. E.	1966	Indirect measurements of atmospheric temperature profiles from satellites: I. Introduction. *Mon. Wea. Rev.*, **94,** 351–362